DOCUMENT RESUME

ED 152 807 TM 006 968

AUTHOR Wissman, Dale J.
TITLE Audio-Visual Proficiency Testing: Annotated Bibliography. Interim Report for Period October 1976-June 1977.
INSTITUTION Air Force Human Resources Lab., Brooks AFB, Texas.
REPORT NO AFHRL-TR-77-54
PUB DATE Oct 77
NOTE 18p.
AVAILABLE FROM Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402 (Stock Number 771-122/72)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Annotated Bibliographies; *Auditory Visual Tests; Elementary Secondary Education; *Equivalency Tests; Job Skills; Job Training; Literature Reviews; Performance Criteria; Post Secondary Education; *Testing; Test Validity; *Visual Measures

ABSTRACT

This bibliography examines the major works on audio-visual testing of job proficiency, evaluates what has been done, and determines whether the development of an audio-visual proficiency test is feasible. Research in audio-visual testing began in the midforties with film slides and synchronized recorded sound. During the early 1950's sound motion pictures were used in industry. Since then, simple video presentations of standard printed tests have been frequently used. However, the majority of research lacks a thorough investigation evaluated against a valid performance criterion. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

AFHRL-TR-77-54

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: TM IR

In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

AIR FORCE HUMAN RESOURCES LABORATORY

ED152807

TM006 968

AUDIO-VISUAL PROFICIENCY TESTING: ANNOTATED BIBLIOGRAPHY

By

Dale J. Wissman, Sgt, USAF

PERSONNEL RESEARCH DIVISION
Brooks Air Force Base, Texas 78235

October 1977
Interim Report for Period October 1976 – June 1977

Approved for public release; distribution unlimited.

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235

NOTICE

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This final report was submitted by Personnel Research Division, under project 7719, with HQ Air Force Human Resources Laboratory (AFSC), Brooks Air Force Base, Texas 78235.

This report has been reviewed and cleared for open publication and/or public release by the appropriate Office of Information (OI) in accordance with AFR 190-17 and DoDD 5230.9. There is no objection to unlimited distribution of this report to the public at large, or by DDC to the National Technical Information Service (NTIS).

This technical report has been reviewed and approved for publication.

LELAND D. BROKAW, Technical Director
Personnel Research Division

DAN D. FULGHAM, Colonel, USAF
Commander

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFHRL-TR-77-54	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): AUDIO-VISUAL PROFICIENCY TESTING: ANNOTATED BIBLIOGRAPHY	5. TYPE OF REPORT & PERIOD COVERED: Interim, October 1976 – June 1977
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Dale J. Wissman	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Personnel Research Division, Air Force Human Resources Laboratory, Brooks Air Force Base, Texas 78235	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 62703F, 2313T608
11. CONTROLLING OFFICE NAME AND ADDRESS: HQ Air Force Human Resources Laboratory (AFSC), Brooks Air Force Base, Texas 78235	12. REPORT DATE: October 1977
	13. NUMBER OF PAGES: 16
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

audio-visual testing
motion picture testing
personnel selection
proficiency testing
test administration

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report contains an annotation of published reports dealing with the use of audio-visual media for proficiency testing. Reports included discuss investigation of audio-visual administrations of standardized printed tests as well as motion picture filmed tests assessing job knowledge and skill. Reports dealing with the use of audio-visual media in teaching were not included, though references to several bibliographies dealing with the use of motion pictures in education are listed. The reviewed research covers a period from 1945 to 1976.

The majority of the investigations of motion picture filmed tests were done between 1945 and 1958. Most of the more recent research is limited to the audio-visual display of printed tests during test administration.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

Item 20 (Continued)

Although the published results of the reviewed reports vary from extremely positive to quite negative, the general finding of this review stresses the paucity of thorough investigations using the audio-visual medium in proficiency testing.

PREFACE

The work reported in this study was accomplished under project 2313T6, Force Acquisition, Assignment, and Evaluation.

AUDIO-VISUAL PROFICIENCY TESTING: ANNOTATED BIBLIOGRAPHY

INTRODUCTION

The purpose of this bibliography is to examine the major works on audio-visual testing of job proficiency and to condense them into one source to evaluate what has been done in this area and to determine whether the development of an audio-visual proficiency test is feasible, practical, and cost effective. This document will also provide the reader with a convenient reference as to what has been accomplished in the use of audio-visual materials for the purpose of testing aptitude and proficiency. The bibliography was compiled from a review of research studies, dissertations, and other investigations involving the use of audio-visual media for testing job proficiency. The research covers a span from 1941 to the present. All potential sources identified by the author were consulted. No pertinent article was purposefully omitted.

It has often been demonstrated that audio-visual aids such as training films, educational television, and motion pictures are an invaluable asset in any educational curriculum (Dale, 1969; Harcleroad, 1962; Schram, 1962; Wendt & Butts, 1962). A good deal of evidence from research supports the conclusion that "Properly prepared audio-visual materials can help us teach our subjects with increasing effectiveness at all levels of learning" (Dale, 1969, p. 140). Audio-visual films stimulate motivation to learn because they appeal to sight and sound sense modalities in a more complete, involving manner. A student will get completely interested in a film presentation of data, while a written verbal description of the same topic may lack meaning or seem dull and dry. Audio-visual presentations appeal to students of varied intellectual abilities. Films of complex procedures instruct not only the student who reads and writes well, but also the pupil who is not verbally gifted. Procedures and events that are actually seen, whether physically or via films, are better recalled and understood. Complicated interactions among parts requiring intricate operations can be demonstrated and identified. Written descriptions of these procedures often seem ambiguous and confusing.

Scientific investigation of learning via an audio-visual medium has stimulated a great deal of research. This bibliography has focused on investigations of testing using audio-visual media; thus, most studies involving the use of audio-visual techniques in education are not reviewed here. The reader is referred to McClusky (1950); Lumsdaine (1953); May and Lumsdaine (1958); and Hsia (1968) for a detailed bibliography of audio-visual learning references. One particular report by Kendler, Kendler, and Cook (1951) merits separate reference. In this investigation, the authors compared the implications of stimulus-response (S—R) learning theory with the design of audio-visual learning aids. This investigation outlines a series of experiments which demonstrate how the primary postulates of S—R theory—drive, cue, response, and reward—are identified and systematically varied within an audio-visual learning situation. Investigations such as this one are especially valuable because they build practical, demonstrable applications on accepted theoretical foundations.

Research in audio-visual testing began in the midforties with film slides and synchronized recorded sound. In 1945, Thelen used this methodology to evaluate "overt responses which can be used for valid prediction of behaviors assumed to constitute the goals of education" (p. 35). In October 1943, an Army organization, the Psychological Test Film Unit, was established at Santa Ana Army Air Base, Santa Ana, California, as part of the Aviation Psychology Program. Its primary purpose was to develop and extend the work already begun on an experimental program of motion picture test construction and on allied problems involved in the psychological use of films (Gibson, 1947). The removal of the preflight school from Santa Ana Army Air Base and the later termination of large-scale pilot training programs resulted in the reduction of experimental aptitude and proficiency test development including audio-visual research.

During the early 1950's, the Instructional Film Research Program at The Pennsylvania State University, in conjunction with the Personnel Research Branch, The Adjutant General's Office, developed research of uses of sound motion pictures in industry (Carpenter, Greenhill, Hittenger, McCoy, McIntyre, Murnin, & Watkins, 1954; McIntyre, 1954). The major paper to come out of this research reports the most successful application of audio-visual methods to proficiency testing of all reviewed research. At about this same time, the Air Force Personnel and Training Research Center performed a series of research studies of radar aiming point identification motion picture tests (Church, 1957; Herman & Church, 1954). The

purpose of this research was to develop motion picture tests in the task of aircraft observer during bombardment. Following these investigations, the military was strangely silent on audio-visual research until the 1970's. Investigators at the Air Force Human Resources Laboratory published a series of reports in 1974 concerning the development of devices used to measure the training success and promotion potential of maintenance personnel (Shriver, Hays, & Hufhand, 1974). The fourth volume in this series reports an unsuccessful effort to use video media as an approach to performance testing. The authors' recommendation that video should not be further considered as a testing medium for performance testing curtailed additional military research in this area.

In the years between the radar aiming point investigations and the maintenance personnel performance investigations, the area of audio-visual proficiency testing was dominated by educational researchers. During the years from the late 1950's to the mid-1970's, many studies applying clinical and projective tests to audio-visual situations were developed. The major conclusion of these studies showed that a projected administration of a standardized test is as functional as a printed administration.

This audio-visual testing methodology, frequently used to economically test large groups, employs the simple video presentation of a standard printed test on a movie screen or through closed circuit or educational television. A wide range of objective and projective tests including the Peabody Picture Vocabulary Test (Bart, 1971; Fargo, Crowell, Noyes, Fuchigami, Gordon, & Dunn-Rankin, 1967), a delinquency proneness scale, (Curtis, King, & Kropp, 1963), and the Terman Concept Mastery Test (Curtis & Kropp, 1973), have been adapted for audio-visual administration. Nearly all authors report similar results using either the printed or the audio-visual medium. Differences between the studies arise in the estimation of relative costs of printed versus audio-visual test administration and the selective advantages of either of these media for low ability or low socioeconomic groups (Bart, 1971). The majority of the authors listed the advantages of an audio-visual test administration as increased subject motivation, control of pacing of test items, economy in time and personnel, and increased consistency and reliability between repeated administrations.

The audio-visual testing methodology utilizing actual filmed sequences of occupational tasks demonstrating techniques and identifying areas of improper procedure has received much less research to demonstrate its effectiveness. The majority of this research was done by the military shortly after World War II. The results of this research are relatively difficult to evaluate because, for most of these investigations, no criterion against which to establish validity correlations exists. The most used criterion, final course grade or pass/fail, is useful for aptitude testing but not for proficiency testing. A criterion based on scores on a written proficiency measure would not be appropriate because such a test would be correlated to a large degree with verbal skills. One of the major reasons for using audio-visual tests is that they aren't influenced to a large degree by verbal ability. A fundamental problem of any testing in an audio-visual medium then becomes to find a criterion against which to measure test validity or to devise some other way of demonstrating the validity and appropriateness of the device.

Although a validity criterion has not been precisely defined for the areas evaluated by audio-visual testing, several investigations have been made, and the investigators have reached some interesting conclusions. The feasibility of developing sound motion picture tests which yield a high reliability has been demonstrated. These tests have been shown to have a relatively high correlation with paper-and-pencil tests. In Motion Picture Testing and Research, Gibson (1947) summarizes his findings by saying:

> It is likely that there are types of human aptitudes and ability, only touched upon by the tests described, which cannot be adequately measured by the relatively static problems and questions presented by ordinary test methods but which can be demanded by setting up tasks arising from the continuous flow of events portrayed on the motion picture screen. (p. 98)

The majority of the research summarized in this review points to the striking lack of a thorough investigation of an audio-visual test using a concrete procedural task evaluated against a valid performance criterion that is relatively independent of verbal skills.

Index

		Page
1.	Applegate, N.J. *Test and evaluation of criterion testing in the format of discrete motion pictures.* CNET Support Report 4-76. Pensacola FL: Chief of Naval Education and Training Support, June 1976	6
2.	Bart, L.E. A comparison of the effectiveness of televised and conventional administrations of objective scales. *Dissertation Abstracts International*, 1971, 32B, 2980–2981	6
3.	Carpenter, C.R., Greenhill, L.P., Hittinger, W.F., McCoy, E.P., McIntyre, C.J., Mumin, J.A., & Watkins, R.W. The development of a sound motion picture proficiency test. *Personnel Psychology*, 1954, 7, 509–523	6
4.	Church, S.A. *Refinement and validation of an aiming point identification motion-picture group test.* AFPTRC-TN-57-142. San Antonio TX: Air Force Personnel and Training Research Center, December 1957	7
5.	Curtis, H.A., King, F.J., & Kropp, R.P. Validity studies of scores from a delinquency proneness scale. *Psychological Reports*, 1963, 12, 271–278	8
6.	Curtis, H.A., & Kropp, R.P. Standard and visual administrations of the concept mastery test. *Audio-Visual Communications Review*, 1973, 10, 38–42	8
7.	Fargo, G.A., Crowell, D.C., Noyes, M.H., Fuchigami, R.Y., Gordon, J.M., & Dunn-Rankin, P. Comparability of group television and individual administration of the Peabody Picture Vocabulary Test. *Journal of Educational Psychology*, 1967, 58, 137–140	8
8.	Gibson, J.J. (Ed.) *Motion pictures testing and research.* Report #7, AD-651 783. Army Air Force Aviation Psychology Program Research Reports, 1947	9
9.	Herkowitz, J. Filmed test to assess elementary school-aged children's perception of embedded figures which appear to move away from stationary backgrounds. *Dissertation Abstracts International*, 1971, 32A, 3075–3076	10
10.	Herman, I.L., & Church, S.A. *Analysis of radar aiming point identification motion picture group tests.* AFPTRC-TR-54-2. San Antonio TX: Air Force Personnel and Training Research Center, April 1954	10
11.	Hopkins, K.D., Lefever, D.W., & Hopkins, B.R. TV vs. teacher administration of standardized tests: Comparability of scores. *Journal of Educational Measurement*, 1967, 4, 35–40	11
12.	McIntyre, C.J. Sex, age, and iconicity as factors in projective film tests. *Journal of Consulting Psychology*, 1954, 18, 337–343	11
13.	Shriver, E.L., Hayes, J.F., & Hufhand, W.R. *Evaluating maintenance performance: A video approach to symbolic testing of electronics maintenance tasks.* AFHRL-TR-74-57(IV). Dayton OH: Advanced Systems Division, Air Force Human Resources Laboratory, July 1974	12
14.	Stoller, R.J., & Geertsman, R.H. Construction of a final examination to assess clinical judgment in psychiatry. *Journal of Medical Education*, 1958, 33, 837–840	12
15.	Tennis, M.H. A comparison of an audio-visual test with a written test. *Florida Journal of Educational Research*, 1970, 12, 109–117	13
16.	Thelen, H.A. Testing by means of film slides with synchronized recorded sound. *Educational and Psychological Measurement*, 1945, 5, 33–48	13

1. Applegate, N.J. *Test and evaluation of criterion testing in the format of discrete motion pictures.* CNET Support Report 4-76. Pensacola FL: Chief of Naval Education and Training Support, June 1976.

Criterion testing in the format of a discrete motion picture was tested and evaluated. Two different recruit groups operating as control (N = 92) and experimental (N = 100) groups were used to test and evaluate the use of pre- and post-training film examination questions testing learning of a Navy training film, Oxygen Breathing Apparatus. Major questions to be answered by this report involved differences in student scores by testing medium and comparative production and administrative costs of the two methods of testing. The control group members were given a printed (paper-and-pencil) pre- and post-test while the experimental group members were administered the same pre- and post-test questions by means of motion picture film.

T-tests of significance indicate that there appears to be no difference (at the .05 level) in test scores between student groups that can be accounted for in the mode of testing (paper or film). An economic analysis of relative production costs indicates that incorporating test questions in training film formats tends to be prohibitive and favors printed questionnaires as the more cost efficient method to be used. A serendipitous finding revealed that learning gains made by using a structured instructional film were quite outstanding when comparing pre- and post-test differences.

2. Bart, L.E. A comparison of the effectiveness of televised and conventional administrations of objective scales. *Dissertation Abstracts International*, 1971, 32B, 2980–2981.

The principal objective of this dissertation research was the "investigation of the effect of an audio-visual method of test presentation, television, on disadvantaged students. Its aim was to determine whether the mode of test presentation influenced the results and, if so, to learn whether an audio-visual presentation would affect one socioeconomic group more than another." Two hundred forty third grade students representing two socioeconomic groups participated in the study; 120 were disadvantaged students and 120 were middle-class students. Every student received TV and teacher administrations of three tests: the Colored Progressive Matrices, the Columbia Mental Maturity Scale, and the Peabody Picture Vocabulary Test.

Three 3-way analyses of variance and subsequent t-tests were used to analyze the data. The results revealed that the middle-class group scored significantly higher than the disadvantaged group on all three tests. The findings also revealed that the television administration resulted in higher scores than did teacher-administered tests of the three instruments. The hypothesis that the difference between socioeconomic groups would be decreased significantly as a result of a televised administration was found to be true only for the Peabody Picture Vocabulary Test. Both socioeconomic groups benefited by the television administration for the other two tests. In addition to the advantages of economy and standardization, TV administration serves to reduce the verbal factor which long has penalized lower socioeconomic groups.

3. Carpenter, C.R., Greenhill, L.P., Hittinger, W.F., McCoy, E.P., McIntyre, C.J., Murnin, J.A., & Watkins, R.W. The development of a sound motion picture proficiency test. *Personnel Psychology*, 1954, 7, 509–523.

The purpose of this research was to determine the feasibility of producing and using sound motion pictures as a means of proficiency testing. Several advantages to motion picture testing listed by the authors are:

1. Action and movement can be realistically presented in films. Most performance or work requires perception of the performer's actions in relation to his job or perception of and adjustments to actions of other persons or operating machines. The full range of actions and movements varying from simple to complex can be presented to test populations by this medium.

2. Sequences of events can be presented in which the spatial-time elements of performance are effectively shown and tested.

3. Motion picture testing allows the concrete-specific presentation of an actual, realistic situation. When a verbal item is read, the subject must "visualize" the situation, then derive the solution from this subjective visualization. Film presentation offers a concrete, uniform situation without the necessity of a subjective interpretation.

4. Sound can be added to the test situation to further add authenticity to the representation of the actual work situation. Job performance often is not highly correlated with verbal ability. Individuals in performance fields may be good mechanics, truck drivers, or repairmen without being able to read well. Sound motion pictures offer a means of evaluating the performance of each individual by offering non-verbal cues or by emphasizing realistic sounds that occur within a work situation.

5. Time spent on each item as well as for the test as a whole is held constant for all subjects. The exposure time factor in film can be varied and made an integral part of the film test.

Any method of testing has its own inherent limitations. Sound motion picture tests also have disadvantages when compared with other materials. Complicated skills and expensive equipment are necessary to produce and distribute a motion picture test. There is a limited range of types of items and methods of scoring for audio-visual tests. There is relative difficulty in changing items after the test has been produced.

The Track Vehicle Repairman course was selected as the course for which the experimental film test was developed. Selection of this course was made based on the fact that much overt, gross behavior as well as fine motor skills and sound cues are crucial to performance. A pool of 200 multiple-choice problems were developed which placed emphasis on the ability to diagnose and correct malfunctioning, to select correct mechanical procedures, and to recognize and understand the characteristics, functions, and interrelationships of parts. Subjects were 326 graduates of the Track Vehicle Repairman course. The criterion measure, an average of the weekly practical grades, was derived from five graphic scales: (1) Quality of work, (2) Application of classroom principles, (3) Manual dexterity, (4) Selection, use, and care of equipment, and (5) Time spent completing work.

A split-half reliability estimate yielded a reliability of .96. The correlation between the film test and the criterion was found to be .73. The two half-tests correlated .72 and .71, respectively, with the criterion. The final written examination used at the ordnance school was found to correlate .68 with the criterion. A test of the significance of difference between the correlations of .68 and .73 found that this difference could have arisen by chance.

The authors concluded the following:

1. On the basis of the evidence presented, the feasibility of developing sound motion picture tests which yield a very high reliability has been demonstrated.

2. There was little demonstrated difference between this particular film test and the final paper-and-pencil examination in the adequacy with which the criterion was presented. The criterion, however, appeared to favor verbal tests. A major problem for future research will be to evaluate the test against an actual performance criterion. Film tests can be practical to administer, objectively scored, and make it possible to test areas of performance not amenable to paper-pencil testing.

4. Church, S.A. *Refinement and validation of an aiming point identification motion-picture group test.* AFPTRC-TN-57-142. San Antonio TX: Air Force Personnel and Training Research Center, December 1957.

This research is a refinement of the motion picture research, *Analysis of radar aiming point identification motion picture group tests* (AFPTRC-TR-54-2), by I.L. Herman and S.A. Church. The original research under this project is summarized later in this bibliography.

The purposes of the research described in this report were to refine the stimulus presentation and scoring technique and to find out the relationship, if any, between what the tests measured and intermediate validity criteria such as grades and ratings received during training. A motion picture of the radar scope display during a practice bomb run was used for the test. Two forms of the test were administered to 90 rated flying officers attending an observer training course. The reliability of the 38-item aiming point identification motion picture group test was .90, and the equivalent forms reliability was .76. Correlations of .28 and .44 (both significant at the .01 level) with "average score on flight mission" and "instructor's rating of radarscope interpretation," respectively, indicated the validity of the radar aiming point test. It was recommended that this motion picture test be given further evaluation as a predictor of instructor ratings of radarscope interpretive skills.

5. Curtis, H.A., King, F.J., & Kropp, R.P. Validity studies of scores from a delinquency-proneness scale. *Psychological Reports*, 1963, 12, 271–278.

The purpose of this study was to investigate two modes of administration (paper-and-pencil and slide projector) of a delinquency-proneness (D-P) scale in relation to criterion data concerning past and future school and legal difficulties and withdrawal from school. Four hundred White tenth-grade students were administered two versions of a constructed opinion survey. The first version consisted of a 92-item survey on a printed page. The second version was identical except that the items were projected one at a time on a screen using a 35mm slide projector.

The authors list several weaknesses in the sample selection. These were (1) all subjects were White, (2) the population was relatively homogeneous with regard to age, and (3) the policy of the school was to exclude students convicted of serious crime. These deficiencies tended to limit the variance of the sample. So few subjects received adverse ratings on "past legal problems" that little can be written about the relationship of the predictor with it. The analysis of the other variables shows the projected test scores to be generally more predictive than are the present printed test scores except when "future legal problems" is the criterion. The authors conclude that "nothing is to be gained by substituting a projected opinion scale for a printed one."

6. Curtis, H.A., & Kropp, R.P. Standard and visual administrations of the Concept Mastery Test. *Audio-Visual Communications Review*, 1973, 10, 38–42.

Items consisting of word pairs were serially presented one at a time on a television screen in this investigation in which item exposure time was the principal variable being studied. It was hypothesized that televised administration of the Concept Mastery Test would offer audible material and visual imagery as well as operational control of the subject during the testing process. The limited size of the screen prohibited projection of the entire test, so items were presented serially. Serial presentation of items demands pacing of the students' progress on the test. The problem being treated in this report was to determine the effect on test scores caused by a change in the exposure time per test item when those items are projected one by one on a screen. The Terman Concept Mastery Test, Form T, which requires the subject to determine whether each pair of words presented is a pair of synonyms or antonyms, was adapted for use in the study. Graduate students (N = 55) at Florida State University were randomly separated into four exposure time groups.

There were no differences between control and projected means for 3 of the 4 groups. Significant differences occurred only under the highly speeded condition. There seemed to be a slight inverse relationship between normal speed of response and the score obtained. To determine the apparent acceptability of projected tests, the reaction of the subjects was closely observed. The authors believe paced visual tests induced a higher level of motivation among subjects than did the traditional presentation. There is "tentative evidence" that tests of this type when administered under paced, projected conditions can be speeded greatly without appreciably altering test reliability and validity.

7. Fargo, G.A., Crowell, D.C., Noyes, M.H., Fuchigami, R.Y., Gordon, J.M., & Dunn-Rankin, P. Comparability of group television and individual administration of the Peabody Picture Vocabulary Test. *Journal of Educational Psychology*, 1967, 58, 137–140.

This study was conducted to examine the feasibility of adapting the Peabody Picture Vocabulary Test for group administration by means of educational television. The objective was to test the hypothesis that scores obtained in group TV administration would not differ significantly from those obtained in individual administration. The investigators believed that if the two administrations were found to be comparable, the economical group administration could be used to screen children and identify those who need further individual study.

A counterbalanced Treatment X Subjects design was utilized in which half had the group presentation first and half had the individual administration first. Subjects were 126 third-, fourth-, and fifth-grade children selected from the University of Hawaii Elementary School. Individually administered standardized intelligence tests placed these children within an I.Q. range of 91–152 with a mean of 123. Individual presentations were administered to each subject as previously described following the Peabody Picture Vocabulary Test standardized procedures. These procedures were adapted for group TV administration. The adaption included an orientation to the task and an orientation to the use of the answer sheets.

An analysis of variance of the scores obtained yielded a Between Administration F ratio of .75. Since the F ratio was less than the critical value, the variation in the data was attributed to chance.

The apparent comparability in scores obtained under the two types of test administration demonstrates the feasibility of the use of the TV administration of the Peabody Picture Vocabulary Test for group testing. The authors list several important implications in group TV test administration: (1) TV screening is more economical in time and personnel because many subjects can be tested at once by fewer administrators, and (2) taped presentations provide consistency and high reliability between repeated administrations. Educational television has been used successfully for teaching and demonstration purposes. This pilot study demonstrates its use can be extended to evaluation as a group screening medium.

8. Gibson, J.J. (Ed.) *Motion picture testing and research.* Report #7, AD-651-783, Army Air Force Aviation Psychology Program Research Reports, 1947.

The research described in this report originated in the effort to utilize the motion picture medium for purposes of psychological testing and examining in the Army Air Forces (AAF). Research included in the text was conducted by the Psychological Test Film Unit, a continuation of the Perceptual Research Unit of the Psychological Section, Office of the Surgeon, Headquarters, AAF Training Command. Its primary purpose was to develop the work already begun on an experimental program of motion picture test construction and on allied problems involved in the psychological use of films.

The most important research objective of the film unit was the construction of motion picture tests for aircrew classification purposes. The general procedure was to formulate a hypothesis regarding a function thought to be valid for prediction of success in training in one or more of the aircrew specialties. The experimental test was then put together and administered to a group of aviation students in an early phase of their training. The validity of the test was determined by correlating the test scores with success or failure in later phases of aircrew training.

General areas in which tests were constructed are as follows:

Aptitude Tests:

1. Tests of ability to judge motion and locomotion.
2. Tests of ability to judge distance.
3. Tests for spatial orientation.
4. Tests of ability to perceive slight motion.
5. Tests requiring multiple perception.
6. Tests involving sequential perception.
7. Tests of perceptual speed.
8. Tests of comprehension.

Proficiency Tests:

1. Aircraft Recognition
2. Navigation Proficiency
3. Target Identification

Several measures of reliability were taken on the data (Holt, odd-even, first half–second half). Most reliability coefficients ranged from .40 to .75. In most cases validity data were not computed. For a number of tests, the data necessary for computing validities could not be obtained before the termination of large-scale pilot training. The first six aptitude tests completed were validated against graduation-elimination from elementary pilot training. These correlations are, however, in the author's terms "moderate." From the evidence available, both the intercorrelations of motion picture tests, and their correlations with other tests seem in general to be low. The low correlations with other aptitude tests are consistent with the theory that motion pictures are capable of testing functions not amenable to other forms of testing. The generally low intercorrelations between motion pictures themselves indicate uniqueness. The author summarizes his findings by saying,

It is likely that there are types of human aptitude and ability, only touched upon by the tests described, which cannot be adequately measured by the relatively static problems and questions presented by ordinary test methods but which can be demanded by setting up tasks arising from the continuous flow of events portrayed on the motion picture screen. (p. 98)

9. Herkowitz, J. Filmed test to assess elementary school-aged children's perception of embedded figures which appear to move away from stationary backgrounds. *Dissertation Abstracts International*, 1971, 32A, 3075–3076.

The figure-ground perception of elementary school-aged children was evaluated via a 16mm, animated, 20-minute film with sound track. In the test, called the Moving Embedded Figures Test (MEFT), embedded figures appeared to move away from stationary backgrounds. On each of the 27 items comprising the test, the subjects' task was to decide which one of four possible figures was embedded within a background and to indicate his decision as soon as possible by pushing one of four buttons on a box. Performance was evaluated in terms of latency of response.

Eighty school-aged children served as the stratified random sample of subjects. Eight age-sex groups were studied (5–6, 7–8, 9–10, 11–12 years). Two measures of reliability were estimated from a single factor repeated measures ANOV on item latencies for all 80 subjects. The estimated reliability of the mean of the 27 MEFT items was .94. The estimated reliability of a single test was .35.

A two-factor, factorial, ANOCOV indicated a significant age main effect (.01 level), no sex main effect, and no interaction effect. Newman-Keuls Sequential Range Test on adjusted age means indicated significant differences among all age groups (.05 level or better); improvements in MEFT performance paralleling increases in age. Though not statistically significant, at all ages, boys performed better than girls.

One-tailed t-test procedures evidenced that for the total of 80 subjects, for each of the four age groups, and for each of eight age-sex groups, performance on a stationary version of the MEFT was not the same as performance on the regular MEFT (all results were statistically significant at the .01 level or better). Results indicated that the stationary version of the MEFT was a more difficult test than the regular MEFT.

It was concluded that the MEFT was a relatively reliable test measuring figure-ground perceptual ability, appropriate for use with elementary school-aged children.

10. Herman, I.L., & Church, S.A. *Analysis of radar aiming point identification motion picture group tests.* AFPTRC-TR-54-2. San Antonio TX: Air Force Personnel and Training Research Center, April 1954.

The purpose of this research was to analyze and evaluate motion picture tests of proficiency in the task of the aircraft observer bombardment. The development of objective, conveniently administered, reliable, and valid measures of proficiency in the associated skills of a task is an essential factor in improving training and evaluation methods. One associated skill, that of aiming point identification, is easily adaptable to audio-visual testing. Motion pictures of a radar scope during a bomb run offer stimulus material from which a proficiency measure of the observer's task might be obtained. Radar aiming point identification motion picture group tests are composed of motion pictures of a radar scope taken during a bomb run from the initial point to the point of bomb release. While this test could be utilized as a proficiency measure, the current study considered the test as a possible aptitude test. As such, the analysis was primarily concerned with test reliability, item difficulty, and item discriminating power. Also of importance was the identification of test variance in terms of selected printed test variables in the Airman Classification Battery and the test's ability to discriminate between observer and nonobserver subjects. Subjects were 2,330 basic trainees who were completely naive to the observer task.

Hoyt reliability estimates reveal a .91 reliability coefficient for all forms of the test. Though there was a wide range of both difficulty indexes and discrimination indexes, the mean difficulty level was .47 with a standard deviation of .18. Performance on the aiming point tests was correlated with the individual test and stanine scores on the airman classification battery (ACB) in an attempt to identify aiming point test variance in terms of printed test variables. Correlations of the aiming point test with each of the 13 subtests of the ACB vary between .30 and .52. Validity measures were not

available on the aiming point test. Subsequent investigations were planned to investigate the validity of the tests. As part of an earlier study, the test was administered to 75 experienced, rated Air Force officers who had just completed the AN/APQ-24 radar course. A comparison of the naive and experienced subjects indicates that the experienced subjects do significantly better than do the naive subjects. This would appear to indicate that on these tests, selection and training of the experienced aircraft observers made a statistically significant contribution to remembering and locating a point in a pattern of radar returns.

The authors concluded that the aiming point tests do adequately discriminate among untrained subjects. A multiple correlation coefficient of .56 with Dial and Table Reading and Pattern Comprehension indicates that performance on aiming point tests can be predicted rather well by the tests of the ACB. Trained subjects perform significantly better than do untrained subjects.

Further research conducted within aircrew training schools and on the job to develop tests of proficiency as well as aptitude was recommended.

11. Hopkins, K.D., Lefever, D.W., & Hopkins, B.R. TV vs. teacher administration of standardized tests: Comparability of scores. *Journal of Educational Measurement*, 1967, 4, 35–40.

This investigation of the relative comparability of scores achieved on two different administrations of an elementary science test emphasizes the importance of the term "standardization." The authors stress the idea that the term "standardization" implies control of certain test conditions. When not controlled, these conditions may have a significant effect on test variance. Listed as variables to be controlled are size of group being tested, familiarity of the examiner with the examinee, and test environment.

Fifth- and sixth-grade students from 20 schools were randomly assigned by school to one of two groups, closed-circuit TV administration or teacher administration. All variables were controlled as nearly as possible. Both groups of students had prior experience with educational TV, so novelty factors associated with TV were deemed minimal. The Metropolitan Science Test was administered to the approximately 1,800 students in each group. Analysis of variance (mode-of-administration x class size x sex) was the principal statistical technique employed. The findings from the statistical analyses showed no significant main effect of mode of administration or class size, but a highly significant interaction effect. With TV administration, both grade levels evidenced relatively higher mean scores in the large classes and lower mean scores in the regular-size classes. The authors summarized the differences between administrations as, "The lower performance of the large group with teacher-administration may have reflected greater teacher difficulty in communicating directions to a large group, more examinee problems in learning and following directions, and greater reluctance to ask questions."

12. McIntyre, C.J. Sex, age, and iconicity as factors in projective film tests. *Journal of Consulting Psychology*, 1954, 18, 337–343.

This study was undertaken to investigate the probable effectiveness of sound motion pictures as a medium for projective personality testing and to attempt to define some of the characteristics which should be built into such a test. It was hypothesized that since a major determinant of the success of projective tests is the extent to which subjects interpret the stimuli as projections of their own personalities, filmed projective tests will enhance this projection and add to the subjects' responses. Also hypothesized was that the variables of age, sex, and iconicity (realness or lifelikeness) would cause more or less projection. Projection was defined as the degree to which the subjects' perception of the protagonist agrees with their perception of themselves as measured by the Minnesota Multiphasic Personality Inventory (MMPI). Five experimental scenes were developed from Thematic Apperception Test (TAT) cards. Subjects were 425 college students in elementary courses in education, psychology, and sociology.

Iconicity was not found to have any significant effect on projection as measured. The degree of projection in each item did not significantly vary between filmed or printed versions. The hypotheses of age and sex affecting projection were not confirmed. While the results did not permit valid conclusions to be drawn about the relative effectiveness of films per se, the experimenter offers a few subjective impressions. A motion picture is realistic partly because it depicts people behaving. The more a person is shown behaving, the more a situation is defined; i.e., it loses ambiguity. By

definition, the TAT stimulus must be ambiguous enough to allow the subject to project his perception of the situation. To this extent, a film test following the TAT paradigm may not be the most effective approach for this purpose.

13. Schriver, E.L., Hayes, J.F., & Hufhand, W.R. *Evaluating maintenance performance: A video approach to symbolic testing of electronics maintenance tasks.* AFHRL-TR-74-57(IV), AD-A005 297. Dayton OH: Advanced Systems Division, Air Force Human Resources Laboratory, July 1974.

This volume reports the continuation of an effort to examine methods for simulating the electronics maintenance task as a means of measuring the proficiency of individual technicians. To overcome shortcomings of verbal and symbolic tests, video tape recordings were investigated as a testing medium. Several tests were constructed of varying proficiency areas. One test consisted of a film recording of a task being carried out. The test subject is shown the film and asked whether it was correct and whether proper tools and procedures had been utilized. Another test requires the subject to watch the results of a filmed system checkout with various faults inserted into the system. The subject must then conclude whether the equipment is operating properly and, if not, what is the trouble.

Test administrations of both the video and actual performance tests were not given to any technician due to the difficulties in getting a satisfactory version of the video tests. The reported results are based on administration of just the video materials. "Based upon the results obtained in the individual test areas, it was concluded that video has several inherent characteristics that make it undesirable as a medium for administering performance tests in electronics maintenance."

Specific deficiencies are summarized:

1. The presentation time required is excessive.
2. Subject cannot control or alter sequence of action.
3. Subject becomes bored watching a familiar operation.
4. Many jobs require reference to technical documentation before proceeding with task.
5. Costs of video material development are excessive.

The major drawbacks to video proficiency testing of the electronic technician's job is the lack of flexibility in viewing the test situation. The electronic technician must draw from many sources to diagnose and solve problems of maintenance and troubleshooting. These sources cannot adequately be demonstrated on film.

14. Stoller, R.J., & Geertsman, R.H. Construction of a final examination to assess clinical judgment in psychiatry. *Journal of Medical Education*, 1958, 33, 837–840.

The senior clerkship in psychiatry at the U.C.L.A. School of Medicine is primarily designed to teach students to observe, understand, and clinically evaluate patients with severe emotional illnesses. Administrators have been unsatisfied with standard methods of examination because of their inadequacies in assessing clinical skills. The problem was to develop methods for more adequately assessing these clinical skills in psychiatry. An adequate method should (a) provide uniform conditions of assessment; (b) utilize a uniform, objective criterion against which to measure proficiency; and (c) test clinical skills without introducing irrelevant skills.

Thirty-minute psychiatric interviews were filmed of two patients with different psychopathologies. Five principal clerkship instructors, all psychologists, viewed the films and separately made clinical evaluations of the patients by assigning ratings from 0 to 6 to each of some 300 statements. The statements had been previously selected to represent a general population of statements which could be used for the psychiatric description of any major type of emotional illness. About 100 of these statements which formed the rating criterion were rated sufficiently alike by all the instructors to be represented on a final exam. Individual instructor evaluations were correlated with the criterion evaluations. These correlations ranged from .83 to .94 with a mean of approximately .91. For their final examination, 47 senior medical students were shown the two filmed interviews and made the evaluations with the preselected statements.

The correlation between the ratings of a student evaluation and the ratings of the criterion evaluation for a patient would indicate the correctness of his clinical judgment. No display of students' scores was offered by the authors.

This method of evaluation provides uniform examination conditions and an objective, uniform criterion with which to assess clinical judgment. With modification, this method of assessment is potentially applicable to teaching, evaluation, and research problems in medical fields other than psychiatry. The filmed interview can be replaced by visual, auditory, or tactual stimuli to which clinical judgments can be applied.

15. Tennis, M.H. A comparison of an audio-visual test with a written test. *Florida Journal of Educational Research*, 1970, 12, 109–117.

This report reviews research in the field of audio-visual testing beginning with Thelen in 1945. Basic conclusions drawn from research reviewed by Tennis are as follows:

1. Audio-visual testing could improve measurement of lower grades of behavior (Thelen).

2. Motion picture testing "could save time and money;" also, it is advantageous in the testing of complex situations (Carpenter et al., 1954).

3. Audio-visual testing using television as a testing medium is comparable to conventional testing with added advantages of a higher degree of motivation and more control over administration (Curtis et al., 1973).

The author describes the basic goal of audio-visual testing research as developing more valid, reliable, vivid, and realistic testing procedures. The purpose of the paper was to make a comparison between Thelen's sound-slide test and a comparable written test. Tennis (1970) wanted to find out what the advantages and disadvantages of audio-visual testing are, if students are motivated to a higher degree by audio-visual testing, and if audio-visual testing is more efficient at measuring some behaviors than others.

Two tests were constructed to measure ability to apply principles of elementary science. This behavior was sampled through 15 problem-situations which required subjects to recognize a scientific principle in the solution of a problem. The items of the test were in the multiple-choice, short answer, and true-false format. The audio-visual test consisted of a series of film slides projected on a movie screen. The presentation of the slides was paced to narration describing the problem and authentic sound effects. Subjects were students in grades 5–10 at the University of Chicago Laboratory School. At each grade level, half the students received the audio-visual test first and the written test two weeks later. The other half of the students received the tests in reverse order.

Higher medians were obtained at all grades on the audio-visual test than on the written test, regardless of the order in which the tests were taken. A chi square 2 x 2 contingency table showed different median scores for initial vs. final administrations of either form of the test as well as written vs. audio-visual administration for certain grades. For a few of the grades, the difference between correlation coefficients did not reach significance (.05) although they were in the direction of higher audio-visual scores.

16. Thelen, H.A. Testing by means of film slides with synchronized recorded sound. *Educational and Psychological Measurement*, 1945, 5, 33–48.

Thelen defines evaluation as an attempt "to put the student into situations likely to result in experiences engendering overt responses which can be used for valid prediction of behaviors assumed to constitute the goals of education." Limitations of paper-and-pencil tests are summarized as follows: paper-and-pencil tests "present artificial situations to which the range of kinds of response is limited, and that facility in manipulation of verbal symbols is an important factor which masks to some unknown degree the nonreading abilities to be measured." The present study investigated the possibilities of the sound-slide medium for reducing the loading of verbal symbolism and increasing the participation of students in testing situations. A test film strip was developed to measure "ability to apply scientific principles." The test items were taken from fifth-grade physical science tests. Students in grades 5, 7, 8, and 10 were subjects in the analysis of the constructed test.

As hypothesized, median score increased by grade level. No tests of significance were performed on the data. No validity correlations or test reliabilities were taken. The author summarizes the

advantages of filmed tests as (1) increased uniformity of administration of the test from group to group, (2) higher motivation of the students, (3) minimization of the verbal element with increased validity of testing certain objectives, (4) possibility of appraisal of some fairly sophisticated objectives at low-grade levels. The "realness" of the test situations is greater than with paper-and-pencil tests. Consequently, it should enable more valid predictions as to the behavior of students in similar "real" situations, and this type of prediction is assumed to be the most legitimate purpose of achievement testing.

REFERENCES

Dale, E. *Audio-visual methods in teaching* (3rd ed.). New York: Holt, Rinehart, & Winston, Inc., 1969.

Harcleroad, F. Theoretical formulations in audio-visual communications. *Review of Educational Research*, 1962, 32, 119–126.

Hsia, H.J. On channel effectiveness. *AV Communication Review*, 1968, 16, 245–267.

Kendler, T.S., Kendler, H.H., & Cook, J.C. *Implications of learning theory for the design of audio-visual aids.* HRRL Memo #12. Washington, D.C.: Human Resources Research Laboratories, November 1951.

Lumsdaine, A.A. Audio-visual research in the U.S. Air Force. *AV Communication Review*, 1953, 1, 76–90.

May, M.A., & Lumsdaine, A.A. *Learning from films*. New Haven CT: Yale University Press, 1958.

McClusky, F.D. *The A-V bibliography*. Dubuque IA: Wm. C. Brown, 1950.

Schramm, W. Learning from instructional television. *Review of Educational Research*, 1962, 32(2), 156–165.

Wendt, P.R., & Butts, G.K. Audiovisual materials. *Review of Educational Research*, 1962, 32(2), 141–155.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1977—771-122/72